

Review Article

AI, Misinformation, and the Maternal Health Information Ecosystem: A Narrative Review and Implementation Framework for Sub-Saharan Africa

Adeyemi Sallam Febilola¹, Muhammed Adewale Alimi¹, Isaac Tobechukwu Onah¹, Chikeluba Joseph Emeonye¹, Ahmad Onikoko Abdulgafar², Nyimeawiya Ngema Nyee³, Ibrahim Adekunle Babawale⁴

¹Faculty of Pharmaceutical Sciences, Usmanu Danfodiyo University Sokoto, Sokoto State, Nigeria

²School of Medical Laboratory Science, Usmanu Danfodiyo University, Sokoto (UDUS), Nigeria

³Nursing, Lublin Academia WSEI, Warsaw, Poland

⁴Department of Anatomy, Ladoke Akintola University of Technology, Ogbomosho, Oyo State, Nigeria

Received: Sep 26, 2026

Accepted: Oct 26, 2026

Corresponding author's email:

adeyanju8998@gmail.com

Citation: Febilola AS, Alimi MA, Onah IT, Emeonye CJ, Abdulgafar AO, Nyee NN, et al. AI, Misinformation, and the Maternal Health Information Ecosystem: A Narrative Review and Implementation Framework for Sub-Saharan Africa. *Epidemiology and Health Data Insights*. 2026;2(5):ehdi063. <https://doi.org/10.63946/ehdi/19706>

Copyright: By the author(s).

License: A non-exclusive license by the publisher. Published by Australasia Publishing Group LLP.

Open Access: This article is an open access article distributed under the terms and conditions of the CC-BY Creative Commons Attribution license <https://creativecommons.org/licenses/by/4.0>



ABSTRACT

Background: Generative AI and large language models are changing how people seek health information, but their effect on maternal health in sub-Saharan Africa is unclear. This narrative review synthesizes evidence on generative AI and maternal health misinformation, examines community health workers (CHWs) as information intermediaries, and proposes a conceptual framework for safe AI implementation.

Methods: We ran a purposive search of PubMed/MEDLINE, Web of Science, Scopus, ACM Digital Library, and Google Scholar for publications from January 2014 to August 2026. We synthesized findings narratively and used the Consolidated Framework for Implementation Research (CFIR) to organize implementation determinants.

Results: We found no study of generative AI maternal health misinformation in sub-Saharan Africa. Evidence from high-income settings and non-maternal topics shows AI answers on reproductive health can be inaccurate or inconsistent, and fluent AI text can seem more compelling than human writing. Multi-specialty African benchmarks and a global benchmark review point to a gap between benchmark performance and clinical safety, but maternal-specific performance is unreported. CHWs have a well-documented record of improving maternal and child health, but their ability to identify AI errors is untested. African digital health implementations highlight user-centered design, interoperability, and sustained investment.

Conclusions: We propose, as hypotheses for evaluation, localized benchmarking, retrieval-augmented generation restricted to national guidelines with added safeguards, CHW curricula that include AI literacy, and decentralized research hubs. Unvalidated AI could bypass CHW authority and affect trust, though this is unstudied. Priorities include measuring AI use among pregnant women and prospectively evaluating CHW-facing tools.

Keywords: Generative AI; Maternal Health; Misinformation; Community Health Workers; Sub-Saharan Africa; Implementation Science

Introduction

The global maternal health information ecosystem is undergoing a structural transformation driven by the proliferation of consumer-facing generative artificial intelligence technologies. Historically, pregnant women in Low- and Middle-Income Countries (LMICs) have navigated health information from traditional birth attendants, community elders, radio broadcasts, and formalized Community Health Workers [1,2]. Over the past two decades, this ecosystem expanded to include mobile health applications, internet search engines, and social media platforms [3]. However, the introduction and rapid global scaling of generative AI and large language models represent a shift in how health-seeking behavior is enacted, interpreted, and resolved [4].

Unlike traditional search engines that return hierarchical lists of external links for user evaluation, generative AI tools deliver centralized, conversational, and highly confident answers often stripped of clear primary source verification [5,6]. These models possess a distinct capability to closely mimic the linguistic register of authoritative medical literature, generating a "fluency illusion" that may increase the perceived credibility and persuasiveness of inaccurate data in other settings [7,8]. Consequently, plausible-sounding but clinically false guidance regarding obstetric danger signs, self-managed medication abortion protocols, or pharmacological safety during gestation can be delivered with the same syntactic polish as evidence-based clinical consensus [4,9].

In high-income settings, generative AI for health queries has rapidly transitioned from a digital novelty to a routine element of patient information-seeking behavior [10,11]. While clinicians are beginning to report the operational need to correct AI-derived misinformation during routine consultations [10], the implications of this shift may be greater in LMICs. Across sub-Saharan Africa, structural inequities and fragile information pipelines may increase the risks of relying on unverified digital health information, with potential consequences for maternal and neonatal health [12,13]. Addressing this problem is not only a technical alignment task; it is also a health systems challenge that may affect clinical trust [14,15].

Community Health Workers (CHWs) have historically served as the foundational trust layer between formal health systems and marginalized or vulnerable populations in LMICs [16]. Their efficacy in improving antenatal screening uptake, facility-based delivery rates, and post-natal care compliance across resource-constrained landscapes is well documented [2,17]. However, this robust evidence base was established entirely prior to the contemporary wave of

consumer-facing generative AI deployment [18,19]. Whether CHWs can effectively adapt to fulfill their role as trusted information intermediaries within an increasingly AI-saturated environment is a central question for global health policy [20,21].

Direct data on the use of generative AI by pregnant women in sub-Saharan Africa are lacking. Mobile health applications, search engines, and social media are already part of the information ecosystem [3], so exposure to generative AI is plausible, but we found no study that measures which tools women use, in which languages, on which devices, or at what cost. We therefore treat exposure to AI-generated maternal health information as a plausible but unmeasured scenario. This review addresses two related but distinct problems. The first is consumer exposure: pregnant women who consult publicly available chatbots outside the health system. The second is health-system implementation: the deployment of AI decision-support tools for CHWs. These problems involve different populations, responsible actors, and interventions. We review evidence on both, identify which problem each finding or recommendation concerns, and apply the implementation framework only to the second, where an implementing organization exists.

This narrative review has three objectives: first, to synthesize evidence on AI-generated maternal health misinformation, with attention to sub-Saharan Africa; second, to examine the role of CHWs as information intermediaries in the digital age; and third, to propose a conceptual implementation framework for safe, equitable AI deployment based on the Consolidated Framework for Implementation Research [22,25]. In this review, sub-Saharan Africa is the geographical focus. We use "LMICs" only for literature that is not specific to the region, and "Africa" only for sources that cover countries beyond sub-Saharan Africa, such as the Pan-African AfriMed-QA dataset.

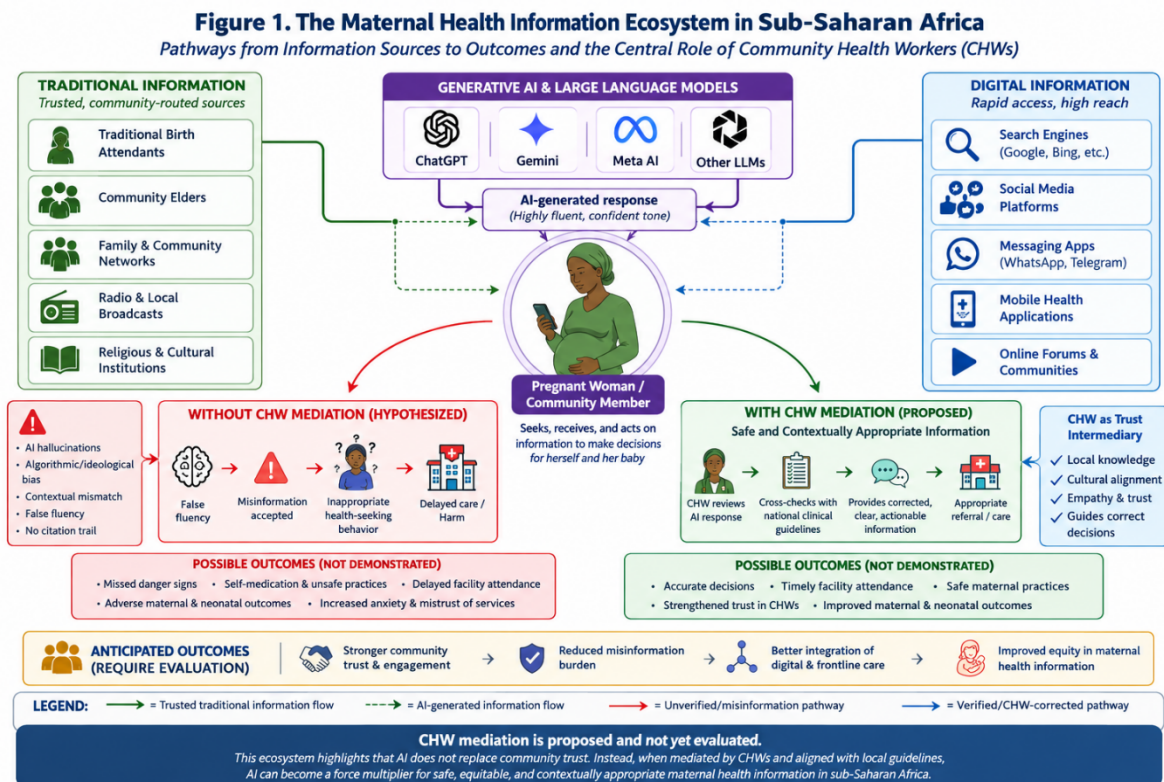


Figure 1. Proposed maternal health information ecosystem in sub-Saharan Africa. The figure shows hypothesized information flows among traditional, digital and generative AI sources, and a proposed corrective role for CHWs. The pathways and outcomes are conceptual and have not been empirically evaluated.

Methods

This study is a narrative review with a purposive literature search. We chose this design because the research question spans computer science, public health, implementation science, and behavioral sociology, and the emergent, heterogeneous evidence base cannot be meaningfully pooled through meta-analysis. We structured and reported the review following the Scale for the Assessment of Narrative Review Articles (SANRA) [23]. It was not conducted as a systematic or scoping review.

We searched PubMed/MEDLINE, Web of Science, Scopus, ACM Digital Library, and Google Scholar between June and August 2026 for publications from January 2014 to August 2026. The complete PubMed/MEDLINE search expression was: ("generative AI"[tiab] OR "large language model"[tiab] OR "ChatGPT"[tiab] OR "artificial intelligence"[tiab] OR "machine learning"[tiab]) AND ("maternal health"[tiab] OR "reproductive health"[tiab] OR obstetric[tiab] OR antenatal[tiab] OR prenatal[tiab]) AND ("LMIC"[tiab] OR "low- and middle-income country"[tiab] OR "sub-Saharan Africa"[tiab] OR "South Asia"[tiab] OR "community health worker"[tiab] OR "resource-limited"[tiab] OR "resource-constrained"[tiab]) AND ("2014/01/01"[dp] : "2026/08/31"[dp]). We also hand-searched the reference

lists of key sources. Only English-language publications were considered. The start date of January 2014 captures the expansion of mobile and digital health in LMICs that preceded generative AI, while the literature on generative AI itself is concentrated after 2022. A small number of foundational sources published before 2014 (on CHW effectiveness, CHW scale-up, information credibility, and the original CFIR publication) were identified through reference-list searching and are cited for foundational context.

Because the search was purposive, the authors selected sources in three steps. First, we screened titles and abstracts for relevance to the three review objectives. Second, the full text of potentially relevant sources was read. Third, a source was retained if it provided evidence on generative AI accuracy, bias or misinformation in health; on CHW effectiveness; or on digital health implementation in African health systems, or if it supplied policy or framework context. When several sources made the same point, we preferred peer-reviewed studies with transparent methods and sources set in sub-Saharan Africa. We do not report screening counts because this was not a systematic process. Search terms included other LMIC regions to capture wider literature, but the evidence retained is predominantly from sub-Saharan Africa.

We did not perform a formal risk-of-bias appraisal, given the heterogeneity of the sources. Instead, we distinguished three kinds of support for each claim: direct empirical evidence on AI performance or misinformation, evidence extrapolated from high-income settings or non-maternal topics, and contextual literature.

We synthesized findings narratively and organized them thematically around the three review objectives. To build the implementation framework, we

mapped the themes onto the five domains of the updated Consolidated Framework for Implementation Research (CFIR 2022) [25]: Innovation, Outer Setting, Inner Setting, Individuals, and Implementation Process. The authors carried out this mapping, using prior work on applying CFIR in LMICs as a guide [24], without stakeholder consultation or a consensus method. The resulting framework is a conceptual proposal and has not been empirically validated.

Results

The Proliferation of Generative AI and the Mechanics of Misinformation

Large language models are not databases of factual information; they are probabilistic engines trained on vast corpora of internet text to predict the most statistically likely next token in a sequence [5]. This architecture has implications for health information seeking. When a pregnant woman queries a standard search engine, the engine retrieves existing webpages for user evaluation (Figure 2). The user must exercise health literacy to differentiate between a peer-reviewed medical portal and an unmoderated community forum [26]. Generative AI can remove this step of source evaluation by synthesizing a direct answer in the voice of an omniscient authority [5,6,11].

Because these models are optimized for fluency, grammar, and confident conversational tone, they may be perceived as credible [6]. Studies indicate that AI-generated text can closely mimic the register of authoritative medical writing, which may increase the persuasiveness of false claims [7,20]. In classroom settings, fluent AI answers can create an illusion of reliability [8]. Whether this applies to maternal health information has not been tested. This dynamic is particularly concerning in maternal health, where reproductive topics are frequently stigmatized, highly personal, and legally restricted, driving women to rely disproportionately on private digital channels [13,27].

Evidence on generative AI and reproductive health is emerging. It points to errors and potential harms, but we found no study that demonstrates harm to pregnant women in LMICs. A systematic review reports that generative AI can accelerate the volume and apparent authority of health misinformation, in what has been described as a "synthetic infodemic" [4]. One analysis reports that models can surface non-vetted, ideologically driven hotlines as neutral resources and fail to flag unscientific medical interventions [28]; this finding rests on a single source and needs replication. Commentaries warn that deploying biased algorithms and unverified text

models in maternal health care could widen existing disparities in resource-limited settings [12], but this is a projected risk and not a measured outcome. Studies of consumer symptom checkers, which predate generative AI, show that these tools cannot take a full medical history or accept flexible symptom input [29]. These limitations may also apply to text-prediction tools used for triage, although this has not been tested in maternal care.

These studies indicate that AI models can generate inaccurate or inconsistent reproductive health information. They do not show how often this happens or whether it harms pregnant women. None of these studies was conducted in, or drew samples from, sub-Saharan Africa, South Asia, or Latin America. Reviews of generative AI and health misinformation draw mainly on evidence from high-income settings, which leaves a geographical gap [40]. This is an important gap in the evidence for the equitable deployment of global health technology.

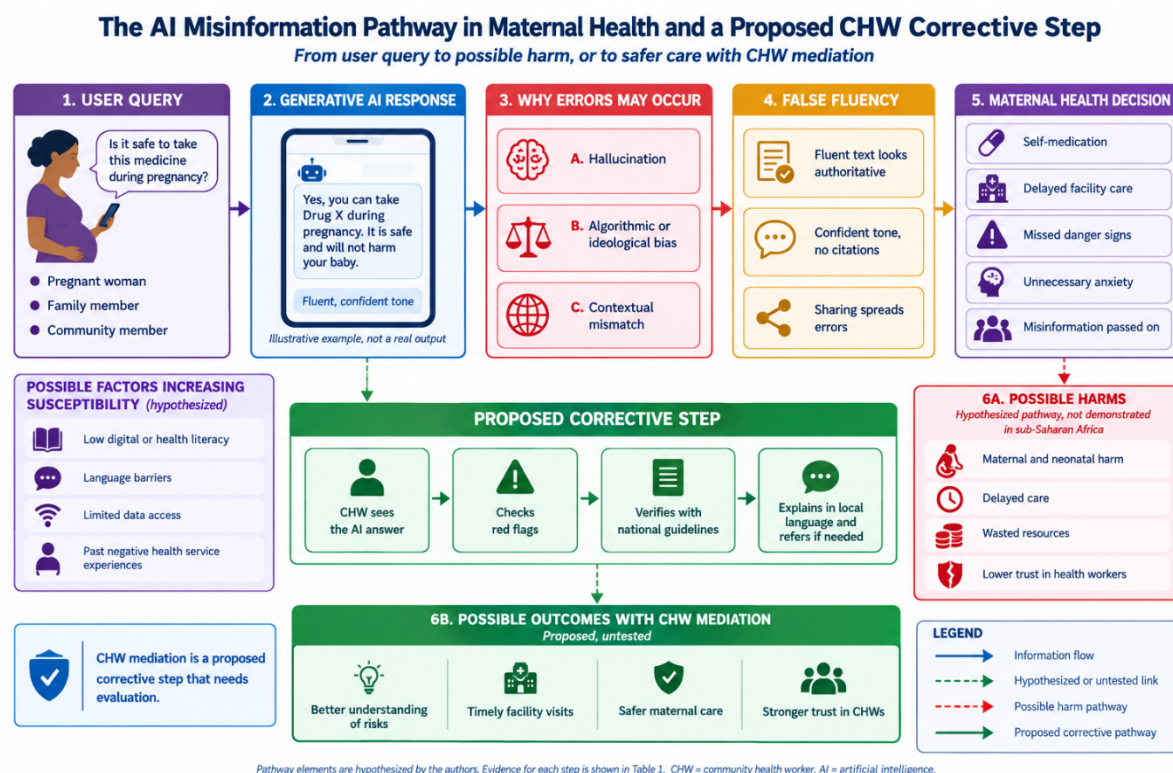


Figure 2. Proposed pathway from a pregnant woman's query to a maternal health decision, with a possible CHW corrective step. Elements are hypothesized by the authors; Table 1 shows the evidence for each step. CHW mediation has not been evaluated

Table 1: Evidence on Errors and Potential Harms of Generative AI in Maternal and Reproductive Health

Category	Evidence	Potential implications (not measured)	Evidence type	Ref.
Inaccurate details on medication abortion	In a comparison of ChatGPT 3.5 and Google Bard on 14 medication abortion questions, ChatGPT scored a mean of 5.26 of 6 for accuracy and 2.57 of 3 for completeness, and Bard 4.5 and 2.14, with no significant difference. Both gave some inaccurate details, such as the oral route for misoprostol and early home pregnancy testing.	Wrong details on dosing route or follow-up testing could lead to unsafe self-management.	Direct (generative AI, reproductive health, no LMIC users)	[9]
Overstated risk in self-managed medication abortion	A qualitative analysis of 65 ChatGPT responses found that the chatbot described self-managed medication abortion as more dangerous than the evidence indicates. The authors interpret this as a risk of perpetuating stigma and affecting reproductive autonomy; they did not measure these outcomes.	Could discourage use of safe care.	Direct (generative AI, reproductive health, no LMIC users)	[49]
Ideologically routed resources	One ethical and legal analysis reports that models can surface non-vetted, ideologically driven resources as neutral advice.	Could direct vulnerable users to biased resources.	Contextual (single source)	[28]
Gaps against clinical guidelines	A benchmarking study against European cervical cancer guidelines and a review of 39 clinical benchmarks report factual gaps and a	Missing citation trails may limit users' ability to verify answers.	Direct for [30] (non-maternal);	[30,31]

Category	Evidence	Potential implications (not measured)	Evidence type	Ref.
	gap between knowledge and practice performance.		contextual for [31]	
Triage limits of symptom checkers (pre-generative AI)	A user study describes people carrying the burden of reasoning from non-specific symptoms.	Text-prediction triage tools might share these limits.	Extrapolated (pre-generative AI)	[29]

Evidence type: Direct = an empirical study of generative AI output or user behavior. Extrapolated = evidence from other AI tools, topics or settings, applied here by inference. Contextual = a review, commentary or analysis. "Direct" does not mean evidence of clinical harm: no entry measured patient outcomes.

The Evidence Gap in Sub-Saharan Africa and the Inadequacy of High-Income Country Benchmarks

Large language models are trained predominantly on digital corpora reflecting High-Income Country health systems, English-language clinical terminology, and Western epidemiological patterns [34]. Medical knowledge and cultural concepts of health, however, are heavily localized; clinical guidelines, endemic disease prevalence, available pharmaceutical formulations, and healthcare behaviors vary dramatically across regions [35]. When AI models optimized for Western clinical settings are deployed in African contexts, they may exhibit algorithmic bias [36]. For example, algorithms trained on predominantly white clinical images fail to accurately recognize dermatological manifestations of diseases on darker skin tones [37]. In maternal health, a model might recommend clinical interventions based on specialized infrastructure that simply does not exist in rural LMIC environments [38]. As a hypothetical illustration, a pregnant woman in rural Nigeria who asks an AI about pre-eclampsia symptoms could receive advice optimized for a different clinical context. This scenario has not been tested.

To assess this disparity and support accountability of model developers, researchers have spearheaded the development of regionally contextualized medical question-answering benchmarks. The most prominent is AfriMed-QA, a Pan-African, multi-specialty medical QA dataset comprising 15,000 questions from more than 60 medical

schools in 16 African countries to evaluate model utility in local environments [39]. When state-of-the-art LLMs, which routinely achieve passing scores on Western standardized clinical exams, are evaluated on clinical benchmarks, a systematic review of 39 benchmarks reports a gap between knowledge and practice performance [31]. AfriMed-QA is multi-specialty and not specific to maternal health, and maternal-specific performance on it has not been reported separately.

The AfriMed-QA evaluation of 30 LLMs reported significant variation in performance across specialties and geographies, and multiple-choice performance that lagged the USMLE-based MedQA benchmark [39]. On the MEDEC benchmark, which uses US clinical notes, recent LLMs performed well at detecting and correcting medical errors but were outperformed by physicians [41]. Neither benchmark is specific to maternal health, and MEDEC is not specific to African settings. Strong benchmark scores do not guarantee safe clinical use [31].

The absence of evidence specific to sub-Saharan Africa is an important gap that research should address. A scoping review of medical AI in low-resource settings describes challenges and strategies for deployment, including the concentration of evidence in high-resource settings [40]. Unvalidated systems may reproduce or widen existing disparities, and consumer-facing text tools could compromise patient safety without local validation [36,40], but this has not been measured in maternal care.

Table 2: Benchmarks and Evaluation Studies of LLM Performance

Benchmark	Focus Area	Scope	Geographic Coverage	Key Finding	Ref.
AfriMed-QA	Multi-specialty medical question answering (English)	About 15,000 questions: 4,000+ expert multiple-choice, 1,200+ short-answer,	Pan-African: 60+ medical schools, 16 countries, 32 specialties	Evaluated 30 LLMs. Performance varied significantly across specialties and geographies, multiple-choice performance lagged the USMLE-based MedQA, biomedical models	[39]

Benchmark	Focus Area	Scope	Geographic Coverage	Key Finding	Ref.
		10,000 consumer queries		underperformed general models, and smaller models struggled to pass. Not specific to maternal health	
Gong et al.	Systematic review of clinical LLM benchmarks	39 benchmarks	Global	Reports a gap between knowledge and practice performance	[31]
MEDEC	Medical error detection and correction in clinical notes	3,848 clinical texts, including 488 notes from three US hospital systems	United States	Recent LLMs performed well but were outperformed by physicians	[41]

Community Health Workers as Essential Trust Intermediaries

In the face of systemic digital misinformation and algorithmic bias, CHWs are a potential point of intervention [17]. In many LMICs, Community Health Workers (CHWs) and traditional birth attendants have long functioned as the primary trusted interface between formal health systems and pregnant women in under-resourced or remote settings [1]. CHWs are deeply embedded within the communities they serve, operating not merely as task-shifted extensions of clinical care but as crucial cultural translators [16].

Evidence for CHW effectiveness in reproductive, maternal, newborn, and child health is substantial. A systematic review reports that CHW-delivered interventions improved several maternal and child health outcomes [2]. Their success is predicated on geographic proximity, linguistic alignment, and a deep, inter-generational reservoir of community trust [16]. CHWs may act as a buffer against rumor-based misinformation, and a CHW may be a trusted source when a woman receives conflicting advice [16]. No study we identified has examined this role for AI-generated information.

However, the vast majority of literature on CHW efficacy was developed prior to the current wave of consumer-facing generative AI. Existing frameworks document the CHW as a trust layer relative to localized word-of-mouth rumors or traditional beliefs [2,17]. They do not address a scenario in which a pregnant woman has already consulted an AI chatbot, possibly on a shared smartphone, before a CHW visit. This scenario is plausible but has not been studied. Emerging digital ecosystem assessments in countries like Nigeria highlight that while structural healthcare stresses persist, rapid smartphone adoption may be driving users toward informal digital channels [18].

Programs that combine community-level trust with mobile health tools may be well placed to observe how AI-generated information affects maternal health-seeking, although no such observations have been reported [3,13]. We propose that CHWs be considered as a possible layer for monitoring and responding to digital misinformation. This is a proposed role that has not been evaluated. Any such role must account for CHW training, workload, supervision, and access to clinical support, and CHW involvement alone cannot guarantee that women receive accurate information.

Table 3: Community Health Worker Digital Health Interventions: Evidence from Sub-Saharan Africa

Country	Intervention	Key Outcome	Lesson Learned	Reference
Rwanda	Community health worker-led digital integrated disease screening system (not maternal-specific)	Streamlined CHW workflows; reduced unnecessary facility referrals; optimized point-of-care data transmission.	Digitizing clinical protocols can strengthen primary healthcare delivery when designed with user input and local integration.	[42]
Ethiopia, Ghana, & Zimbabwe	Primary healthcare digital health interventions (e.g., electronic health tracking)	Cross-country insights revealed variations in data completeness; system fragmentation reduced user trust.	System interoperability and alignment with local health worker workflows are essential for long-term intervention scaling.	[43]

Country	Intervention	Key Outcome	Lesson Learned	Reference
Uganda	Technology Acceptance Model (TAM) digital study	Behavioral intention to adopt digital health tools was high; however, actual daily usage was heavily constrained by structural realities.	Structural barriers (device access, data costs, electricity) dominate over behavioral intention; offline capabilities are essential.	[19]

These interventions did not use generative AI. Findings are extrapolated to AI tools. The Rwandan system was an integrated disease screening tool and not specific to maternal health

Digital Health in Sub-Saharan Africa: Threats and Opportunities

As digital health technologies permeate LMIC health systems, their impact on the community health workforce can be considered in terms of two contrasting possibilities: a threat of administrative displacement and systemic mismatch, or an opportunity for cognitive assistance and health system strengthening.

The Threat: Pilotitis and the mHealth Graveyard

Skepticism among frontline health workers about digital solutions imposed by external actors is well documented among frontline health workers regarding digital solutions imposed by external actors [17]. Scoping reviews tracking medical AI deployment highlight that the vast majority of clinical trials for generative models are conducted exclusively within high-resource settings, leaving an evidence gap regarding their safety and utility in low-resource environments [40]. The implementation history of mobile health in sub-Saharan Africa includes many short-term, donor-funded pilots that were abandoned after funding ended, a pattern described as 'pilotitis' [44].

Unvalidated AI deployment poses multiple interconnected threats to patient safety and institutional integrity. If an unvetted text model provides highly persuasive advice that conflicts with a CHW’s clinical guidance, particularly when based on mismatched high-income country guidelines, it could undermine the authority of local clinicians and affect patient-provider trust, although this has not been studied [28,41]. Furthermore, poorly structured digital tools may add to administrative burden and cause anxiety among workers, and an algorithmic system with a steep learning curve or complex documentation requirements could detract from direct patient care [18,38]. Because consumer-facing language technologies are built on implicit assumptions of continuous high-speed connectivity and seamless cloud compute architectures, they may not function reliably where electricity or data access is unstable [19,34]. In this scenario, AI could amplify misinformation instead

of correcting it. This is a proposed risk and not a measured outcome.

The Opportunity: Digital Health Interventions That Support CHWs

Conversely, a robust counter-current in global health implementation literature advocates for digital tools to serve as assistive cognitive layers that support rather than bypass the community workforce [13]. When frameworks are co-designed with user-input and locally validated, digital health interventions can optimize primary care delivery [42]. Rwanda’s community health worker-led digital integrated disease screening initiative effectively demonstrated how embedding structural clinical protocols could streamline worker workflows, mitigate unnecessary secondary facility referrals, and optimize real-time health system data transmission [42]. Multi-country implementations across nations like Ethiopia, Ghana, and Zimbabwe suggest that digital health architectures can enhance point-of-care tracking, provided that structural fragmentation and data duplication across systems are actively minimized [43].

The operational translation of these interventions depends heavily on user-centric design principles and the components of the Technology Acceptance Model [19]. Empirical evaluations of TAM frameworks among CHWs in resource-constrained environments reveal that while behavioral intention to adopt mobile health tools was high in one Ugandan study, actual daily usage is severely constrained by tangible structural barriers: a lack of smartphone device access, prohibitively expensive data bundles, and unstable grid electricity [19]. Interventions that actively engage local clinical stakeholders from inception and integrate robust offline functionalities are reported to achieve higher long-term acceptance [19,13]. If generative AI frameworks are deployed as locally bounded, Retrieval-Augmented Generation (RAG) architectures, grounded entirely within vetted Ministry of Health guidelines, they may help CHWs check obstetric triage protocols, such as those for postpartum hemorrhage, acting as a clinical aid and not an autonomous decision-maker [46,47]. RAG does not eliminate errors, and this use has not been evaluated in LMIC maternal care. This review concentrates on risks,

and the evidence on the benefits of generative AI in LMIC health systems is still emerging. Digital tools used as assistive layers for CHWs show promise when

they are co-designed and locally validated [13,42]. A balanced assessment would weigh benefits and risks together, and we encourage future reviews to do so.

Table 4: Reported Facilitators and Barriers in Digital Health Implementations in Sub-Saharan Africa

Factor	Reported facilitators (examples: Rwanda, Zimbabwe)	Reported barriers (examples: Ghana, pilotitis)
Design Approach	User-centered, iterative, and co-designed directly with local frontline health workers [42].	Top-down, donor-driven, and built by external developers disconnected from the field landscape [44].
Infrastructure	Deliberate investments in baseline connectivity, offline capabilities, and local charging grids [19,13].	Structural assumptions of constant high-speed internet and infinite cloud-compute capacity [34,40].
Interoperability	Structured integration with national platforms (e.g., DHIS2) via localized open APIs [43].	Isolated, fragmented databases that result in high system friction and severe data duplication [18,43].
Funding Model	Sustained domestic budgetary allocations combined with public-private health partnerships [44].	Short-term donor financing cycles; tool abandonment immediately follows the cessation of funding [44].
Training	Continuous, on-site, iterative training [19]	One-off classroom lectures; no follow-up [19]
CHW Perception	Algorithmic systems are viewed as an assistive clinical co-pilot that strengthens personal capacity [19,42].	Interface is perceived as a tool for intrusive surveillance or an administrative burden [38,43].

Examples come from sources that did not evaluate maternal health or AI outcomes. Cases are grouped by what each source reported, not by a formal outcome criterion

A Conceptual Implementation Framework for CHW-Facing AI in Sub-Saharan Africa

The Consolidated Framework for Implementation Research (CFIR; Figure 3) is widely used to identify barriers and facilitators to implementation [22,25]. It was published in 2009 and updated in 2022 to give more attention to equity, community context, and innovation recipients. The updated framework has five domains: Innovation,

Outer Setting, Inner Setting, Individuals, and Implementation Process [25]. A systematic review of CFIR use in LMICs indicates that its core constructs translate well but need contextual adaptation [24]. CFIR describes how organizations implement an innovation, so we apply it to the health-system deployment of CHW-facing AI tools. It does not address consumer exposure to public chatbots, where no implementing organization exists.

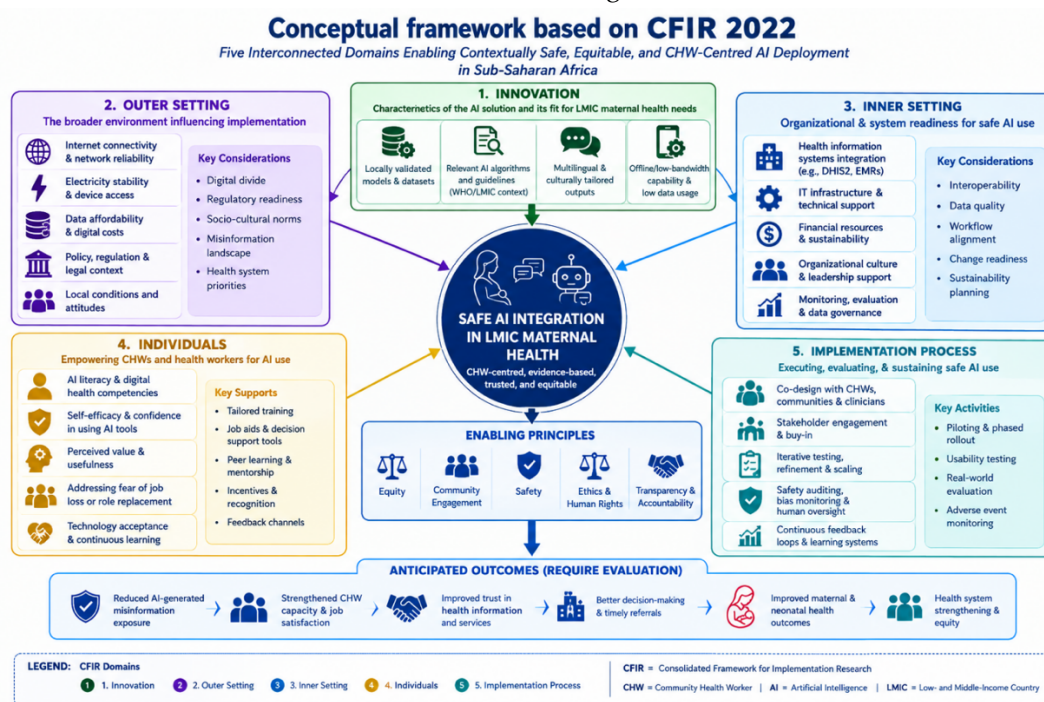


Figure 3. Conceptual framework for CHW-facing AI implementation, based on CFIR 2022 [25]. Domain names follow CFIR 2022 (Innovation, Outer Setting, Inner Setting, Individuals, Implementation Process). The items within each

domain, the enabling principles and the anticipated outcomes are author-proposed and have not been empirically evaluated

Table 5 maps barriers and facilitators from the reviewed literature onto the CFIR 2022 domains. The authors developed this mapping without stakeholder consultation or a consensus method. Construct names follow CFIR 2022. The indicators are author-proposed and are not part of CFIR. The framework is a conceptual basis for future research and has not been empirically evaluated.

Table 5: Proposed Conceptual Framework: CFIR 2022 Domains Applied to AI Implementation in Maternal Health

CFIR Domain	Key Barriers in sub-Saharan Africa Context	Strategic Facilitators	Author-proposed indicators
1.Innovation	Models validated only on Western cohorts; black-box proprietary architectures; high cognitive loads; absent local language tokenizers.	Locally bounded RAG architectures; voice-enabled interfaces in local languages (e.g., Swahili, Hausa, Yoruba); robust offline capabilities	AfriMed-QA clinical accuracy scores; user error detection rates; system latency.
2. Outer Setting	Unstable local grid electricity; high data costs; widespread user exposure to unvetted digital misinformation.	National zero-rating policies for maternal health data; solar power grid partnerships; policies positioning CHWs as trusted arbiters.	Connectivity uptime; local data affordability indices; patient-reported AI information exposure.
3. Inner Setting	Fragmented health databases; short-term donor funding cycles; lack of active API interoperability.	Open API architectures for DHIS2 integration; long-term domestic budgetary streams; dedicated district-level technical support.	System interoperability scores; platform update frequency; maintenance response times.
4. Individuals	Job displacement anxieties; low digital self-efficacy; language barriers; trust misalignment.	Transparent communication framing AI as a supportive clinical co-pilot; continuous peer-led, on-site simulation training.	CHW technology acceptance rates; localized AI literacy scores; post-training self-efficacy metrics
5. Implementation Process	Top-down, donor-driven deployment; absent community feedback loops; no clear reporting channels for hallucinations.	Participatory co-design with CHWs, midwives, and community elders from inception; localized clinical safety audits.	Stakeholder engagement indices; hallucination reporting frequencies; safety audit iteration cycles.

Innovation

This domain assesses specific attributes of the AI technology itself. Innovation Evidence-Base and Innovation Design are plausible determinants of implementation success. If an AI model has only been validated on Western data and fails localized benchmarks like AfriMed-QA, it could introduce clinical risk and may face rejection by local practitioners [31,39]. Innovation Adaptability may be improved by locally bounded RAG designs that draw only on Ministry of Health maternal care protocols [46,47]. Furthermore, Innovation Design must minimize cognitive load, offering intuitive voice-enabled interfaces in local languages that function reliably in low-bandwidth or offline environments [13,40].

Outer Setting

The Outer Setting encompasses the broader economic, political, and socio-cultural context [25]. Local Conditions and Local Attitudes are important. Pregnant women may already be exposed to unvetted digital health advice, which would make CHW-facing

tools relevant, but this exposure has not been measured [26]. Local Conditions, including infrastructure, and Policies & Laws affect feasibility. Unstable electricity and high mobile data costs are reported constraints [18,19]. Strategic facilitators include public-private partnerships for decentralized solar power and policies mandating zero-rated health data streams, ensuring digital tools remain accessible regardless of patient financial status [19,36].

Inner Setting

This domain evaluates the specific organizational structures implementing the intervention [25]. Structural Characteristics, including information technology infrastructure, are reported points of failure; new AI deployments must integrate seamlessly via open API standards with existing national information health software, such as DHIS2, to prevent severe data duplication and system friction [43]. Available Resources requires moving beyond short-term donor cycles; sustainable implementation

necessitates dedicated domestic budgetary allocations and public-private maintenance partnerships for continuous auditing, system updates, and user support [38,44].

Individuals

The ultimate success of digital clinical tools relies on human operators. In CFIR 2022, CHWs are Innovation Deliverers and pregnant women are Innovation Recipients. The Motivation of Innovation Deliverers (CHWs) may affect long-term adoption. If CHWs perceive an algorithmic interface as a surveillance tool or an employment threat, resistance is possible [19]. Implementers must foster psychological safety through transparent onboarding communication, framing AI explicitly as an assistive co-pilot designed to elevate, not replace, local human expertise [45]. Furthermore, worker Capability must be built through continuous, iterative, on-site training rather than isolated classroom lectures, ensuring high confidence when troubleshooting tools during patient visits [19].

Implementation Process

The final domain dictates tactical execution [25]. Engaging stakeholders forms the foundation of user-centered design; top-down deployment models driven entirely by external developers often fail across resource-constrained healthcare landscapes [44]. Facilitators must prioritize Engaging and Planning, including co-design, directly involving CHWs, nurse-midwives, and community leaders from project inception through deployment governance [42,43]. Finally, Reflecting & Evaluating includes establishing continuous clinical safety audits. Localized feedback loops that allow CHWs to instantly flag algorithmic hallucinations ensure the system undergoes continuous refinement based on real-world outcomes. We found no study of automated error detection in CHW-facing tools, which is a research priority.

Proposed Recommendations and Research Priorities

Generative AI is spreading quickly in health contexts, and health systems and regulators should plan for its use. The following recommendations are proposals derived from the reviewed evidence and the authors' judgment. We used no formal consensus method, and none has been evaluated in maternal health settings in sub-Saharan Africa. Recommendations 1 and 2 address consumer exposure (Problem A). Recommendations 3 to 6 address health-system deployment (Problem B). Recommendation 7 addresses both.

- Hold developers of public chatbots accountable (Problem A). Developers and regulators could require transparency and

published safety evaluations of public chatbots on reproductive health queries in relevant languages, with safeguards such as uncertainty signalling and referral to trusted services.

- Invest in public communication and digital literacy (Problem A). Ministries of Health and community organizations could use radio and community channels to explain the limits of AI tools and direct women to trusted care.
- Evaluate models in regionally relevant settings before CHW use (Problem B). Models intended for CHW use should be assessed on regionally relevant benchmarks, such as AfriMed-QA [39], followed by prospective evaluation in local clinical settings. Benchmarks have limits: AfriMed-QA is an English-language, Africa-wide, multi-specialty benchmark, language coverage and data contamination need attention, and benchmark performance does not guarantee clinical safety [31]. A maternal-specific benchmark is an identified need.
- Use retrieval-augmented generation with safeguards (Problem B). For CHW-facing tools, we propose restricting AI to RAG systems grounded in vetted national clinical guidelines [46,47]. RAG can reduce but not eliminate hallucination: retrieval can fail, passages can be misread, and outputs can still contradict the source [48]. We therefore propose additional safeguards: validation and scheduled updating of source documents, explicit signalling of uncertainty, escalation to a qualified clinician for danger signs, prospective safety evaluation against a pre-specified error threshold set by clinical governance bodies, and routine reporting of errors in the field.
- Add AI literacy to CHW training (Problem B). We propose training CHWs on how generative tools work, how to recognize implausible output, and when to refer. Whether such training improves error detection is untested and should be evaluated, with attention to CHW workload, supervision, and access to clinical support. CHW involvement alone cannot guarantee that women receive accurate information.

- Invest in digital infrastructure and interoperability (Problem B). Unreliable electricity and high data costs are reported barriers to digital health adoption in African settings [19], and offline and low-bandwidth designs may reduce, but not remove, this dependence. Infrastructure investment mainly concerns CHW-facing tools in less connected areas, whereas women who already use chatbots are likely to be connected. We propose decentralized solar energy for rural clinics, zero-rated maternal health data, and open API standards for interoperability with systems such as DHIS2.
- Fund locally led AI research hubs (Problems A and B). Relying only on high-income country developers risks technological dependency. We propose redirecting funding toward African NLP researchers, clinical informaticians, and community health organizations to build local models, datasets, and evaluation benchmarks that reflect local languages and epidemiology.

Table 6: Proposed Recommendations for AI in Maternal Health in Sub-Saharan Africa

Recommendation	Target Stakeholders	Implementation phase (indicative)	Proposed indicators	Evidence Base
Developer accountability and platform safeguards	Model developers, regulators, platforms	Phase 1: preparation	Published safety evaluations of public chatbots on reproductive health queries in relevant languages.	[4, 28]
Public communication and digital literacy	Ministries of Health, media, community organizations	Phase 1: preparation	Campaign reach and women's ability to name trusted sources	None identified (research priority)
Regionally relevant evaluation	WHO, national regulators, model developers	Phase 1: preparation	Models for CHW use evaluated on regional benchmarks and prospectively in local settings before approval.	[31,39]
RAG with safeguards	Technology developers, Ministries of Health	Phase 2: pilot and evaluation	Closed RAG grounded in national guidelines, with uncertainty signalling and escalation; error rate measured against a pre-specified threshold set by clinical governance bodies	[46,48]
AI literacy in CHW training	Ministries of Health, training institutions	Phase 1: preparation	Modules piloted and evaluated for effect on error detection	[16,19,21]
Digital infrastructure and interoperability	Donors, governments, telecoms	Phase 3: scale-up	Solar power in rural centers; zero-rated maternal health data; open API integration	[18,19,43]
Locally led research hubs	Donors, academic institutions	Phase 2: pilot and evaluation	Local dataset creation, model tuning, and governance workflows	[34,36,39,40]

Note: Recommendations are author proposals. The sources describe problems or context and do not evaluate the proposed interventions. Sequencing is indicative and is not based on evidence.

Ethics and Governance

Deploying AI in maternal health care raises issues beyond accuracy. These include data protection and consent for sensitive reproductive information, shared phone use that may compromise privacy, gender dynamics in access to devices and in decision-

making, and legal restrictions on abortion that shape both the queries women ask and the responses that are permissible. These issues require explicit governance arrangements and local ethical review before any deployment.

Limitations

This review has several limitations. First, it is a narrative review with a purposive search. We selected sources for their relevance to the review objectives, so the selection is not exhaustive, reflects the authors' judgment, and could not be reproduced exactly by another team. We did not perform a formal risk-of-bias appraisal, and the sources vary widely in design, from benchmark evaluations and qualitative studies to systematic reviews and policy documents. Second, we included English-language publications only, which excludes relevant work in French, Portuguese, and local African languages. We also did not use preprints as evidence, so recent findings in this fast-moving field may be missing.

Third, direct evidence is scarce. No source we identified examined generative AI maternal health misinformation among users in sub-Saharan Africa.

Conclusion

Generative AI is entering the maternal health information environment. Evidence from other settings indicates that its answers on reproductive health topics can be inaccurate or inconsistent, and that fluent text can be persuasive regardless of accuracy. We found no direct evidence of harm to pregnant women in sub-Saharan Africa.

The evidence base in sub-Saharan Africa is sparse. Models trained mainly on high-resource data may perform less reliably where health infrastructure is limited, but maternal-specific performance in these settings has not been measured, and no study we identified has measured how many pregnant women use these tools.

CHWs are trusted intermediaries with strong evidence of effectiveness in maternal and child health.

Evidence on AI errors comes mainly from high-income settings or from multi-specialty benchmarks, and the evidence on CHW effectiveness comes from interventions that did not involve generative AI. Extrapolating these findings to LMIC maternal care is uncertain, and we label extrapolated claims in the text. Fourth, the evidence is predominantly African and the review focuses on sub-Saharan Africa, so findings may not generalize to other LMIC regions. Fifth, AI models change rapidly, and findings on specific models may not apply to current versions. Sixth, the review concentrates on risks and does not examine in depth the potential benefits of AI decision support in LMICs. Finally, the implementation framework was developed by the authors without stakeholder consultation and has not been empirically tested. It should be read as a conceptual basis for future research.

Whether they can identify or correct AI-generated information has not been tested, and workload, training, and supervision would shape any such role. CHW involvement is not a sufficient guarantee of accuracy. Poorly planned AI deployment could also bypass CHW authority or add to administrative burden, although this has not been studied.

If implemented deliberately, guided by implementation science frameworks such as the updated CFIR and evaluated in regional clinical settings, AI could serve as an assistive tool for CHWs. The framework we present is a conceptual basis for future research. Priorities include measuring AI use among pregnant women, evaluating CHW-facing tools prospectively, and building locally led evaluation capacity.

Acknowledgements

Acknowledgments: None applicable.

Ethics approval and consent to participate: Not applicable. This manuscript is a narrative review and does not involve primary data collection from human subjects or animals.

Consent for publication: Not applicable.

Competing interests: The authors declare that they have no competing interests.

Funding: This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Authors' contributions: A.S.F: Conceptualization; study design; methodology; data curation; formal analysis; investigation; visualization; writing – original draft; writing – review and editing. M.A.A: Methodology; data curation; formal analysis;

investigation; writing – original draft; writing – review and editing. C.J.E: Methodology; data curation; formal analysis; investigation; writing – original draft; writing – review and editing. I.T.O: Methodology; validation; investigation; writing – original draft; writing – review and editing. A.O.A: Conceptualization; methodology; supervision; investigation; writing – original draft; writing – review and editing; project administration; corresponding author. N.N.N: Methodology; formal analysis; investigation; writing – original draft; writing – review and editing. B.I.A: Methodology; visualization; investigation; writing – review and editing.

Data availability: All data analyzed during this review are available within the manuscript.

Generative AI statement: The authors declare that generative AI was not used in the creation of this

manuscript, except for visualizations, refinement and grammatical errors corrections.

References

1. Sialubanje C, Massar K, Hamer DH, Ruiter RA. Reasons for home delivery and use of traditional birth attendants in rural Zambia: a qualitative study. *BMC Pregnancy Childbirth*. 2015;15:216. <https://doi.org/10.1186/s12884-015-0652-7>
2. Gilmore B, McAuliffe E. Effectiveness of community health workers delivering preventive interventions for maternal and child health in low- and middle-income countries: a systematic review. *BMC Public Health*. 2013;13:847. <https://doi.org/10.1186/1471-2458-13-847>
3. Feroz A, Perveen S, Aftab W. Role of mHealth applications for improving antenatal and postnatal care in low and middle income countries: a systematic review. *BMC Health Serv Res*. 2017;17:704. <https://doi.org/10.1186/s12913-017-2664-7>
4. Saeidnia HR, Jahani S, Ghiasi N, Keshavarz H. Generative AI and health misinformation: production, propagation, and mitigation: a systematic review. *BMC Public Health*. 2026;26(1):693. <https://doi.org/10.1186/s12889-025-26148-9>
5. Bender EM, Gebru T, McMillan-Major A, Shmitchell S. On the dangers of stochastic parrots: can language models be too big? In: *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*; 2021. p. 610-623. <https://doi.org/10.1145/3442188.3445922>
6. Metzger MJ, Flanagan AJ. Credibility and trust of information in online environments: the use of cognitive heuristics. *J Pragmat*. 2013;59:210-220. <https://doi.org/10.1016/j.pragma.2013.07.012>
7. Spitale G, Biller-Andorno N, Germani F. AI model GPT-3 (dis)informs us better than humans. *Sci Adv*. 2023;9(26):eadh1850. <https://doi.org/10.1126/sciadv.adh1850>
8. Kumar S, Mikayelyan A, Vorfolomeyeva O. Fluency illusion: a review on influence of ChatGPT in classroom settings. *Information*. 2026;17(3):299. <https://doi.org/10.3390/info17030299>
9. Mediboina A, Badam RK, Chodavarapu S. Assessing the accuracy of information on medication abortion: a comparative analysis of ChatGPT and Google Bard AI. *Cureus*. 2024;16(1):e51544. <https://doi.org/10.7759/cureus.51544>
10. Saguil A. Chatbots and large language models in family medicine. *Am Fam Physician*. 2024;109(6):501-502.
11. Huynh AL, Roy TJ, Jackson KN, Lee AG, Liaw W, Hossain MM. Applications of artificial intelligence-based conversational agents in healthcare: a systematic umbrella review. *Int J Med Inform*. 2026;207:106204. <https://doi.org/10.1016/j.ijmedinf.2025.106204>
12. Laguitan RVM, Bernardino GD Jr. Can AI bridge or widen maternal health inequities? *Public Health Chall*. 2025;4(3):e70119. <https://doi.org/10.1002/puh2.70119>
13. Till S, Mkhize M, Farao J, et al. Digital health technologies for maternal and child health in Africa and other low- and middle-income countries: cross-disciplinary scoping review with stakeholder consultation. *J Med Internet Res*. 2023;25:e42161. <https://doi.org/10.2196/42161>
14. World Health Organization. Ethics and governance of artificial intelligence for health: guidance on large multi-modal models. Geneva: World Health Organization; 2024. Available from: <https://www.who.int/publications/i/item/9789240084759>
15. Souvatzi E, Katsikidou M, Arvaniti A, Plakias S, Tsiakiri A, Samakouri M. Trust in healthcare, medical mistrust, and health outcomes in times of health crisis: a narrative review. *Societies*. 2024;14(12):269. <https://doi.org/10.3390/soc14120269>
16. Anstey Watkins J, Griffiths F, Goudge J. Community health workers' efforts to build health system trust in marginalised communities: a qualitative study from South Africa. *BMJ Open*. 2021;11(5):e044065. <https://doi.org/10.1136/bmjopen-2020-044065>
17. Pallas SW, Minhas D, Pérez-Escamilla R, Taylor L, Curry L, Bradley EH. Community health workers in low- and middle-income countries: what do we know about scaling up and

- sustainability? *Am J Public Health*. 2013;103(7):e74-e82. <https://doi.org/10.2105/AJPH.2012.301102>
18. Egwudo AE, Jegede AO, Oyeniyi TA, et al. Integrating digital health technologies into the healthcare system: challenges and opportunities in Nigeria. *PLOS Digit Health*. 2025;4(7):e0000928. <https://doi.org/10.1371/journal.pdig.0000928>
 19. Chraish M, Oyama C, Aoki Y, et al. Bridging the gap between community health workers' digital health acceptance and actual usage in Uganda: exploring key external factors based on technology acceptance model. *PLOS Digit Health*. 2025;4(11):e0001099. <https://doi.org/10.1371/journal.pdig.0001099>
 20. Uzoma C, Chukwuedo G, Isaac E, et al. Chatbots as frontline educators in sexual reproductive health rights: evidence, limitations, and ethical considerations. *Front Digit Health*. 2026;8:1777667. <https://doi.org/10.3389/fdgth.2026.1777667>
 21. Ahsan Z. Integrating artificial intelligence into medical education: a narrative systematic review of current applications, challenges, and future directions. *BMC Med Educ*. 2025;25(1):1187. <https://doi.org/10.1186/s12909-025-07744-0>
 22. Damschroder LJ, Aron DC, Keith RE, Kirsh SR, Alexander JA, Lowery JC. Fostering implementation of health services research findings into practice: a consolidated framework for advancing implementation science. *Implement Sci*. 2009;4:50. <https://doi.org/10.1186/1748-5908-4-50>
 23. Baethge C, Goldbeck-Wood S, Mertens S. SANRA: a scale for the quality assessment of narrative review articles. *Res Integr Peer Rev*. 2019;4:5. <https://doi.org/10.1186/s41073-019-0064-8>
 24. Means AR, Kemp CG, Gwayi-Chore MC, et al. Evaluating and optimizing the consolidated framework for implementation research (CFIR) for use in low- and middle-income countries: a systematic review. *Implement Sci*. 2020;15:17. <https://doi.org/10.1186/s13012-020-0977-0>
 25. Damschroder LJ, Reardon CM, Widerquist MAO, Lowery J. The updated Consolidated Framework for Implementation Research based on user feedback. *Implement Sci*. 2022;17(1):75. <https://doi.org/10.1186/s13012-022-01245-0>
 26. Swire-Thompson B, Lazer D. Public health and online misinformation: challenges and recommendations. *Annu Rev Public Health*. 2020;41:433-451. <https://doi.org/10.1146/annurev-publhealth-040119-094127>
 27. Mlotshwa L, Plüg S, Simuyaba M, et al. Sexual and reproductive health interventions for mobile adolescents and young people in sub-Saharan Africa: a scoping review. *BMC Public Health*. 2025;25(1):3903. <https://doi.org/10.1186/s12889-025-25119-4>
 28. Anokwuru CC, Nwuzoh MI, Eneh S, et al. Regulating algorithmic tools in reproductive health: ethical and legal challenges. *Front Reprod Health*. 2026;8:1771550. <https://doi.org/10.3389/frph.2026.1771550>
 29. You Y, Gui X. Self-diagnosis through AI-enabled chatbot-based symptom checkers: user experiences and design considerations. *AMIA Annu Symp Proc*. 2021;2020:1354-1363.
 30. Pavone M, Innocenzi C, Macellari N, et al. Assessing the accuracy of large language models on European guidelines for cervical cancer: an in silico benchmarking study. *BJOG*. 2026;133(4):771-778. <https://doi.org/10.1111/1471-0528.70095>
 31. Gong EJ, Bang CS, Lee JJ, Baik GH. Knowledge-practice performance gap in clinical large language models: systematic review of 39 benchmarks. *J Med Internet Res*. 2025;27:e84120. <https://doi.org/10.2196/84120>
 32. Abubakar N, Shehu H, Akinyemi MM, Abdulazez HA, Ariyo AA, Godswill JM. AI-generated misinformation and the crisis of digital trust: strategic communication approaches for information integrity. *J Arts Humanit Soc Sci*. 2026;3(2):292-304. <https://doi.org/10.69739/jahss.v3i2.2039>
 33. Bashardoust A, Feuerriegel S, Shrestha YR. Comparing the willingness to share for human-generated vs. AI-generated fake news. *Proc ACM Hum Comput Interact*. 2024;8(CSCW2):1-21. <https://doi.org/10.1145/3687028>
 34. Owoyemi A, Owoyemi J, Osiyemi A, Boyd A. Artificial intelligence for healthcare in Africa. *Front Digit Health*. 2020;2:6. <https://doi.org/10.3389/fdgth.2020.00006>

35. Kpanake L. Cultural concepts of the person and mental health in Africa. *Transcult Psychiatry*. 2018;55(2):198-218. <https://doi.org/10.1177/1363461517749435>
36. López DM, Rico-Olarte C, Blobel B, Hullin C. Challenges and solutions for transforming health ecosystems in low- and middle-income countries through artificial intelligence. *Front Med* (Lausanne). 2022;9:958097. <https://doi.org/10.3389/fmed.2022.958097>
37. Dowie T. Exploring the diagnostic capability of artificial intelligence in dermatology for darker skin tones: a narrative review. *Cureus*. 2025;17(10):e94909. <https://doi.org/10.7759/cureus.94909>
38. Wahl B, Cossy-Gantner A, Germann S, Schwalbe NR. Artificial intelligence (AI) and global health: how can AI contribute to health in resource-poor settings? *BMJ Glob Health*. 2018;3(4):e000798. <https://doi.org/10.1136/bmjgh-2018-000798>
39. Olatunji T, Nimo C, Owodunni AT, et al. AfriMed-QA: a Pan-African, multi-specialty, medical question-answering benchmark dataset. In: *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics*. 2025. p. 1948-1973. <https://doi.org/10.18653/v1/2025.acl-long.96>
40. Al-Ganad A, Al-Shahdhi A, Al-Dhaifi O, Hajeb E, Hajeb H, Al-Motarreb A. Deploying medical AI in low-resource settings: a scoping review of challenges and strategies. *Front Digit Health*. 2026;8:1743634. <https://doi.org/10.3389/fdgth.2026.1743634>
41. Ben Abacha A, Yim WW, Fu Y, Sun Z, Yetisgen M, Xia F, Lin T. MEDEC: a benchmark for medical error detection and correction in clinical notes. In: *Findings of the Association for Computational Linguistics: ACL 2025*. Vienna, Austria: Association for Computational Linguistics; 2025. p. 22539-22550. <https://doi.org/10.18653/v1/2025.findings-acl.1159>
42. Ngabonziza JCS, Suleiman K, Mohamed NS, et al. Implementation and evaluation of a community health workers-led digital integrated diseases screening system to provide healthcare for patients at community level in Rwanda. *EBioMedicine*. 2026;125:106168. <https://doi.org/10.1016/j.ebiom.2026.106168>
43. Simbini T, Adimado E, Adjorlolo S, et al. Digital health interventions in strengthening primary healthcare systems in Sub-Saharan Africa: insights from Ethiopia, Ghana, and Zimbabwe. *PLOS Digit Health*. 2026;5(1):e0000863. <https://doi.org/10.1371/journal.pdig.0000863>
44. Huang F, Blaschke S, Lucas H. Beyond pilotitis: taking digital health interventions to the national level in China and Uganda. *Global Health*. 2017;13(1):49. <https://doi.org/10.1186/s12992-017-0275-z>
45. Goldie J, Dennis S, Hipgrave L, Coleman A. Practitioner perspectives on the uses of generative AI chatbots in mental health care: mixed methods study. *JMIR Hum Factors*. 2025;12:e71065. <https://doi.org/10.2196/71065>
46. Lewis P, Perez E, Piktus A, Petroni F, Karpukhin V, Goyal N, et al. Retrieval-augmented generation for knowledge-intensive NLP tasks. In: *Advances in Neural Information Processing Systems 33*. 2020. p. 9459-9474.
47. Xiong G, Jin Q, Lu Z, Zhang A. Benchmarking retrieval-augmented generation for medicine. In: *Findings of the Association for Computational Linguistics: ACL 2024*. Bangkok, Thailand: Association for Computational Linguistics; 2024. p. 6233-6251. <https://doi.org/10.18653/v1/2024.findings-acl.372>
48. Ji Z, Lee N, Frieske R, Yu T, Su D, Xu Y, et al. Survey of hallucination in natural language generation. *ACM Comput Surv*. 2023;55(12):248. <https://doi.org/10.1145/3571730>
49. McMahon HV, McMahon BD. Automating untruths: ChatGPT, self-managed medication abortion, and the threat of misinformation in a post-Roe world. *Front Digit Health*. 2024;6:1287186. [doi:10.3389/fdgth.2024.1287186](https://doi.org/10.3389/fdgth.2024.1287186)